Volume I: Technical Proposal

BAA Number	N00014-17-S-B001
Proposal Title	**Detecting and Aligning Narrative Variants for Countering Disinformation**
Prime Offeror	School of Computing and Information Sciences Florida International University
ONR Code	341
Subcontracts	None
Type of Organization	Public/State Controlled Institution of Higher Education: Minority Institution (MI) & Hispanic Serving Institution (HSI)
Technical Contact	Professor Mark Finlayson 11200 S.W. 8th Street ECS Building, Room 362 Miami, FL 33199-0001 markaf@fiu.edu (305) 348-7988 (office) (305) 348-3549 (fax)
Administrative Contact	Robert Gutierrez Assistant VP for Research 11200 S.W. 8th Street MARC Building, Room 430 Miami, FL 33199-0001 gutierrr@fiu.edu 305-348-2494 (office) 305-348-4117 (fax)
Total Amount Requested	$400,000 • Base Period: $200,000 • Option Period: $200,000
Type of Instrument	Grant
Period of Performance	September 1, 2017—August 31, 2019 (2 years) • Base: September 1, 2017—August 31, 2018 (1 year) • Option: September 1, 2018—August 31, 2019 (1 year)
Place of Performance	Florida International University Modesto A. Maidique Campus (MMC)

Table of Contents

1. Project Abstract

The increasing volume, variety, and velocity of modern information streams has made it increasingly easy for our adversaries to effectively promote inaccurate, misleading, or outright false information that serves their strategic goals. This "disinformation warfare" has seen strikingly effective use in several recent cases, but perhaps no more notably than in Ukraine's 2014 revolution and Russia's subsequent annexation of Crimea. In that conflict, Russia muddied the information waters effectively enough for long enough--confusing its adversaries, misleading the foreign press, and rallying its own citizens, soldiers, and sympathetic Crimeans--that it was able to create a *fait accompli* land grab without any significant kinetic response.

In this project, we will develop new technology for detecting and analyzing disinformation in information streams. The basic premise of the work is that an adversary will be trying to push its own "narrative" of events to serve its strategic goals, and to do this it will introduce specific narrative variants in the form of alternative events, facts, or statements. These narrative variants may range from a single misleading statement carefully calibrated to sow doubt, to a set of statements that presents an alternative, more favorable interpretation of accepted events, to a completely alternative set of facts that contradict reality. Building on prior, DARPA-funded work on the Analogical Story Merging (ASM) algorithm, we will develop new approaches to detect and analyze these variants. **First**, we will develop methods to classify documents in information streams as story variants: does a specific document contain a variant of the story which we are tracking? **Second**, we will develop techniques for aligning a story variant with related variants, including potentially the "canonical" or "target" narrative that has been separately identified as ground truth. This alignment process will involve identifying the major narrative breakpoints as well as detailed event differences, using an augmentation of the existing ASM approach. **Third**, we will test our approach on several sets of data, which we will annotate for ground truth alignment. We will begin with existing sets of story variants which have already been collected and are well understood. We will also undertake major data collection to collect story variants for geopolitical events over the past two decades, including the Crimean crisis and the Estonian cyber conflict, which we will also annotate and use to test our approach. **Finally**, we will set the stage for future work by developing and applying a classification scheme of communicative strategies to the parts of the variants, and begin building classifiers for assigning disinformation strategies to specific communicative acts: for example, is a statement intended to dismiss key narrators of the target narrative, so as to undermine their authority? Is it intended to distort the underlying evidence of the target narrative, so as to diminish its power? Is it intended to distract from the target narrative by changing the topic? Or perhaps is it intended to dismay potential actors by introducing previously unimagined, negative consequences?

These new approaches and data will be an important step forward to building our capabilities to detect, understand, and counter disinformation warfare. The output of this set of capabilities could take many useful forms: for example, red flag warnings in commander's information centers about disinformation actions and their intent, and automatic recommendations for counter-information to be disseminated in response.

2. Technical Approach and Justification

2.a. Motivating Scenario

Consider the following scenario:

> *In November 2013 Victor Yanukovych, president of Ukraine, suspended the signing of the Ukraine-EU political association and free trade agreement, implying a dramatic pivot away from the EU toward Russia. This resulted in the massive, country-wide Euromaidan protests and a revolution which resulted in parliament removing Yanukovych from office on 21 February 2014. Pro-Russian media claimed that the size of protests had been exaggerated by the foreign press, and were orchestrated by the U.S. and other western powers to pull Ukraine to the west, join NATO, and continue NATO's attempts to encircle Russia. Yanukovych called his removal a "coup d'état" sponsored by foreign powers, and compared the Euromaidan protests to the rise of the Nazi Party. Days after Yanukovych's removal from office and his flight to Russia, pro-Russian demonstrations were held in Sevastopol, Crimea, an autonomous, pro-Russian territory ceded by Russian to Ukraine in 1954, and by 27 February unmarked troops resembling Russian special forces ("little green men") were seen in occupying key positions in the territory, including the Crimean parliament. On 4 March, Vladimir Putin, president of Russia, described these forces as "local self-defense units" and denied any involvement of Russian forces on the ground. Less than two weeks later, after claiming 95% support for annexation in a hastily held referendum, the Crimean parliament formally requested Russia admit the breakaway republic, and on 21 March the Russian Federal Assembly formally annexed Crimea into the Russian Federation. Later reports suggested that at most 30% of Crimeans actually voted for annexation.*

Figure 1: Summary of the 2014 Ukrainian Revolution and the Annexation of Crimea [12], [15], [16]

In this situation, at least two alternative narratives were present. As shown in Figure 2, on the western side was a narrative describing a popular rebellion against an unpopular leader, culminating in a punitive and exploitative action by that leader's allies. On the Russian side was a narrative describing meddling by foreign powers, an illegitimate coup, and an intervention to protect a vulnerable population. Each of these narratives are *variants* of the other, in that they describe the same set of events, have overlapping parts, but differ in important ways. The Russians promoted their variant through official announcements and comments, stories in the sympathetic mass media, and disinformation spread via social media. Each alternative element of their story had a strategic function, for example, to dismiss the legitimacy of western rapporteurs so as to undercut the western narrative, distort the facts so as to support their version of events, and distract the public from Russia's actual motivations and actions [12]. Russia's campaign was chillingly effective, and we must develop methods to detect, understand, and counter this type of disinformation warfare.

In this project we propose to develop new technologies to automatically detect alternative narratives about ongoing geopolitical events (§2.b) and automatically align those narratives to identify their similarities and differences (§2.c), develop reference datasets for evaluating these new approaches (§2.d), and begin to develop ways of automatically classifying the strategic function of those differences so that we may quickly and effectively counter disinformation actions (§2.e). We conclude with an overview of the end-to-end capabilities (§2.f), and a discussion of future naval relevance of the work (§2.g).

Narrative A (Western Media)	Narrative B (Pro-Russian Media)
A1. Leader defies will of people	B1. Leader defies foreign powers
A2. Spontaneous, major protests	B2. Foreign powers sponsor minor protests
A3. Removal of leader	B3. Removal of leader
A4. Resource grab by external actors	B4. Country splits

Figure 2: Outlines of opposing narratives present in the scenario summarized in Figure 1.

2.b. Detecting Narrative Variants

The first problem we will address is detecting when narrative variants—relative to some narrative of interest—are being transmitted. We will call the narrative of interest the *target* narrative. For example, we may be interested in finding variants of the Crimean crisis scenario above, starting from some initial description of the relevant sequences of events. Finding narrative variants is not a trivial problem, as they can be substantially different from the target narrative, and also potentially split across multiple information streams over long periods of time. We will tackle the first part of this problem. The second part—finding and merging large amounts of disparate information across time and space into a narrative—we will leave to future work because it is beyond the scope of resources for this effort.

Thus, for the purposes of this project, we will assume that we have a set of documents containing various narratives on a particular topic (e.g., foreign affairs). The problem is as follows: given a target narrative (e.g., a summary of the Crimean crisis), find all the documents in the set that represent variants of the target. We will approach this as a supervised machine learning (ML) classification problem, where the features will be lexical and semantic information extracted from the stories, such as keywords, tf-idf scores, key events, key characters, and so forth. Importantly, the features must be designed in way such that the ML classifier is parameterizable by the target narrative: the target narrative's features will influence weights of the features for the classifier that then allow narrative variants for that target to be identified.

We will leverage recent work in our lab on building event extractors, character extractors, and other discourse-level analysis methods [2]. We will also leverage cross-document coreference approaches for finding connections between documents [11]. This approach naturally builds upon recent work of ours where we have developed narrative ML classifiers, where we showed we can separate narrative from non-narrative texts with high performance [3].

Because data collection and curation is time-consuming, and geopolitical narrative data will not be immediately available to us in curated form for evaluation, we will do our initial work on this problem on cultural story variants (folktales). Folklorists have already done quite a bit of work in identifying sets of narratives variants: for example, Cox identified several hundred variants of the Cinderella story [1], while Zipes collected more than fifty variants of Little Red Riding Hood [17]. Folklorists have done such analyses for thousands of tales [14], all of which represents an easily collected and used set of variants to test our approach. These can be quickly mixed together to generate test sets of arbitrary size and difficulty to evaluation our technique. Later in the project, when we have collected geopolitical narrative variants sets, we will re-evaluate the technique on that data.

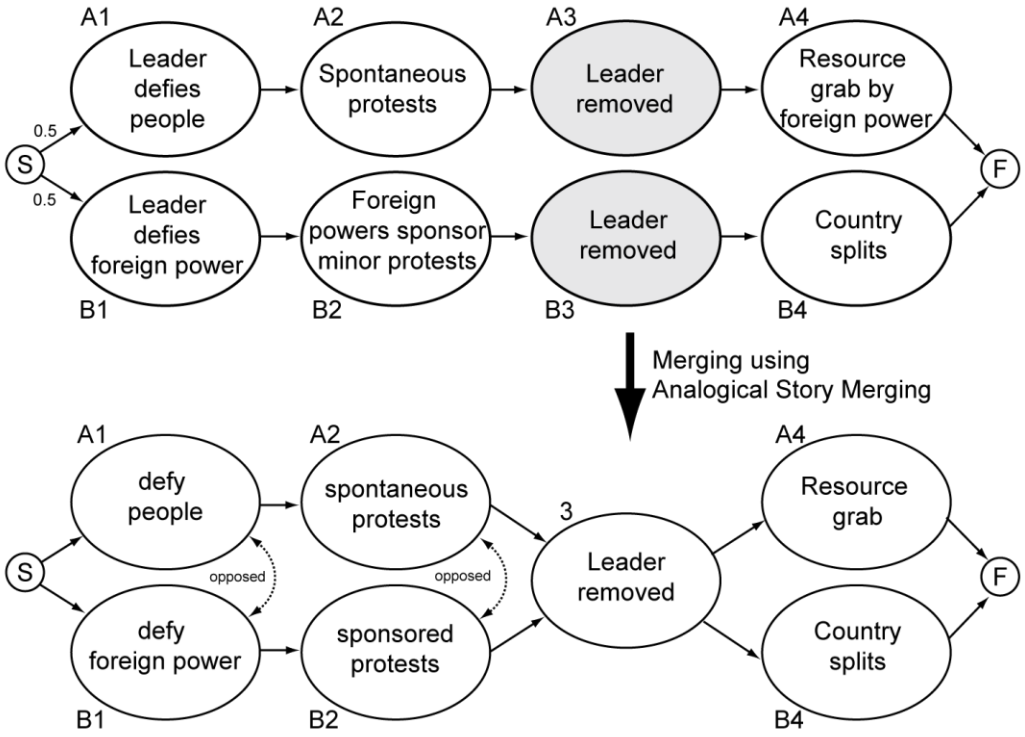

Figure 3: Alignment of the two narrative variants, identifying shared nodes and opposed interpretations of the same events.

2.c. Aligning Narrative Variants

Having identified a set of related narrative variants from a large set of document, the next problem is to understand their similarities and differences. Our approach to this will build on prior work in our lab on automatically extracting the structure of narratives via a technique called *Analogical Story Merging* (ASM) [8], [9]. We have used ASM to automatically extract event sequences from stories and find their similarities and differences. In our prior work, we focused on high-level plot structures, where similarity was expressed in abstract ways (for example, finding all the *villainies* of a set of stories, no matter if they appeared as a *kidnap*, a *murder*, an *insult*, or something else). The problem here is related but different. In this task, we wish to take sets of events extracted from a set of narrative variants and align them in a constrained way, finding common points of intersection determined by reference to the same events in world. This is similar to a sequence alignment problem, which historically has been solved by something like, say, the Needleman-Wunch algorithm [4]. The difference here is that we need to be able to consider multiple events as sub-events of one larger event (for example, one narrative may describe an *leader removed*, while a narrative variant may describe a *demand for resignation* followed by *leader flees*).

Our approach to this alignment problem builds on our prior work. First, we process every narrative document through our standard semantics pipeline, which does co-reference resolution, event detection, semantic role labeling, temporal relationship extraction, and cross-document linking. This produces a set of linear sequence of events, with one sequence per narrative, as shown at the top of Figure 3. The Analogical Story Merging algorithm then uses a specific merge strategy

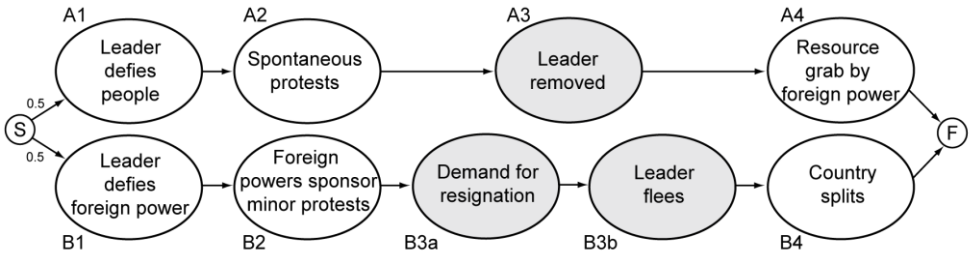

Figure 4: Illustration of the problem of sub-events. The merge model must be able to align A3 with B3a and B3b simultaneously.

tailored for narrative variants to find shared nodes and create a new model which describes the similarities and differences of the variants as a finite state machine. While this is similar to our prior use of ASM, we will need to develop a new merging strategy for similarity detection. In particular, we will use an iterative approach where we start by matching near-identical nodes first, then allow these matches to constrain less exact matches. The key variable is how similar should two nodes be for matching? We will approach this in an experimental way, using hand-annotated alignments of events (which will be generated by trained annotators) to allow us to learn models of the degree of match required.

In the option phase of the project, we will seek to extend this work to tackle the problem of identifying sub-events that comprise larger events, as noted above. This problem is illustrated in Figure 4, where the same event is described in two different variants but at different levels of granularity. Figure 4: Illustration of the problem of sub-events. The merge model must be able to align A3 with B3a and B3b simultaneously.To accomplish this alignment we will leverage knowledge encoded in existing datasets such as Research Cyc (a database of commonsense knowledge encoding event/sub-event relationships [13]), and Framenet (a structured resource that describes the constituent parts of various events and their inter-relationships [5]).

2.d. Curation of New Data

As noted already, we will begin with folklore variants because they are abundant, easily accessible, and well understood. However, our ultimate goal is to apply these techniques to narratives in the geopolitical sphere, and so we will work to collect and curate narrative variants in that space that can be used to test our algorithms. Our first target will be naturally, the Crimean crisis. We will also look at other Russia-related conflicts in the past decade, such as the Estonian cyber conflict of 2007. For these events, we will extract full narratives that express different variants from online resources such as Lexis-Nexus or blogs (via, for example, the Blogtracker data service [10]), using the variant detector developed in the first task of the project. We will also work with experts in disinformation warfare to identify other potential narratives to investigate. We will annotate these sets of narrative variants for alignment in the same way as we annotated the folktales above, using trained annotators to hand-identify alignments of events to produce gold-standard data to evaluate our alignment system.

2.e. Classifying Communication Strategies

In the option phase, we will begin work on using the differences identified by the alignment system to identify disinformation strategies in play. We will develop an annotation scheme starting from Ben Nimmo's "4D" conceptualization of Russia's disinformation tactics: dismiss, distort, distract, and dismay [12]. We will train human annotators to apply these labels to different event interpretations identified in our narrative variants. These labels can then be used in a supervised ML classification approach, to develop a system that can automatically assign one of these tactics to event characterizations. We also allow the possibility that we will find parts of variants that will use multiple tactics simultaneously, or will embody additional tactics not covered in the 4D approach: when we find instances of narrative elements that don't fit the scheme, we will expand it with new categories.

2.f. End-to-End Capability

The work proposed above will enable the following capabilities:

1. Detection of narrative variants, given a target narrative
2. Alignment of those variants with each other, to identify differences
3. Categorization of those differences as to the communicative tactics in play

The output of such a set of a capabilities could take many forms: for example, red flag warnings in commander's information centers about disinformation actions and their intent, and automatic recommendations for counter-information to be disseminated in response.

We will evaluate these technologies first on folktales (because of their abundance and accessibility), and then on new datasets we will collect focusing on conflicting geopolitical narratives, especially those put forth by Russia. This technology will be a significant step forward in our ability to automatically monitor, assess, and quickly react to disinformation actions by adversaries.

2.g. Future Naval Relevance

Knowledge of the theater of operations is increasingly informed and shaped by the communications of ourselves, our allies, neutral parties, and our adversaries. This knowledge is critical not just to the intelligence analyst, but to anyone operating in the theater, from the captain of an aircraft carrier, a navy fighter pilot engaging targets, or marines engaged in combat on the ground. As the Ukraine-Russia conflict has demonstrated, adversaries can use disinformation tactics to achieving information dominance—confusing and distracting us enough to significantly degrade our ability to use kinetic force, and so interfere with our ability to achieve our strategic and tactical goals. The Navy has a deep interest in developing techniques to counter disinformation warfare to maintain the effectiveness of its fighting force.

3. Management Approach

3.a. Personnel

The team will consist of the PI, Dr. Finlayson and two graduate students. Dr. Finlayson will lead the team, directing the students. There will be no subcontractors or subrecipients.

3.b. Facilities

The work described in this proposal will be carried out at the School of Computing and Information Sciences (SCIS) in the Engineering and Computer Science Building (ECS) on the Modesto A. Maidique Campus located in Miami, FL. SCIS provides computing services such as file, compute, web, email, messaging, backup, print, and other computing services. Networking services include a 10 Gigabit Ethernet core network that interconnects rack mounted switches and servers. All school desktop systems are connected by 1 Gigabit switched ports. The network is highly redundant with multiple fiber and copper paths and is designed with routing fail-over capacity. SCIS also provides continuous automated monitoring of its network and servers. The building subscribes to the university's 802.11 WiFi network and SCIS maintains a legacy research WiFi network. The network interconnects at 10GBs to the campus backbone, which provides a 10GBs connection to the NAP of the Americas to provide for connections to the Internet, Internet2, Florida and National Lambda Rail, and CLARA (South American Research) networks.

SCIS computing systems feature a variety of open-source, commercial development, and scientific software products from numerous vendors including IBM, Microsoft, ESRI, MathWorks, etc. SCIS provides middleware technologies to support web services, and the computing environment takes advantage of hundreds of open-source software solutions including Apache with full mods, PHP, Perl, and many others. Many of SCIS's shared infrastructures provide virtualization services.

SCIS maintains a data center incorporating an 8+ node research and instructional Linux Beowulf cluster, multiple virtualization machines, and multiple general compute servers.

SCIS also maintains research and instructional labs, and computer classroom facilities. The school operates seven instructional laboratories for use by undergraduates and graduate students in support of our computer science and information technology degree programs. Our instructional labs offer students access to Windows 7/XP, CentOS Linux, and Mac OS X, which run a variety of software development tools, libraries, databases, and have the capacity to host virtual machines. The school has dedicated servers for student files and computing services and a printer in each lab. Each student receives at least 25GB of backed up file storage space. Students can login remotely to several Linux and Solaris file and compute servers. In addition to workstations provided in the open labs, students may connect their laptops to the SCIS network via WiFi. Students may also use their laptops with 40" LCD displays to collaborate on programming assignments or other joint projects. SCIS also maintains a Visualization Lab in the ECS building, which houses two Sharp PN-L802B 80" Class AQUOS BOARD LED Displays which are multi-touch capable and configured to support high load applications and 3-D rendering.

The SCIS Technology Group maintains all the school's computing facilities (total research and instruction: 26 labs, 350+ desktops, 100+ servers, layer 2 and 3 networking). The SCIS Technology Group consists of 5 FTE of permanent, professional staff assigned to all the school's research and instructional laboratories management. In addition, there are at least 2 FTE of temporary students specifically assigned to laboratory assistance. The SCIS Technology Group staff is organized into three groups: Engineering Services, including Networking, Systems, Desktop, and Help Desk Support; Business Services including Technology Procurement, Asset Management, and Budget/Contract Management; and a Marketing Technology group that promotes the school via digital and social media outlets.

3.c. Government-Furnished Materials

This project assumes no government-furnished equipment, hardware, software, or information.

4. Current and Pending Project and Proposal Submissions

4.a. Current Support: NIH R01

Title of Proposal and Summary: Narratives in the Informational Patient Society and Their Association with Health Behavior

Source and amount of funding: $167,415 direct cost ($230,000 total cost), via subcontract from UCLA. NIH Contract No 5R01GM105033-01; UCLA Subaward # 0160 G SB206.

Percentage effort devoted to each project: 33% summer effort for Co-I Mark Finlayson (1 month/year).

Identity of Prime Offeror and complete list of subwards: UCLA is the prime on this grant, and FIU is the only subcontract.

Technical contacts:

UCLA Technical Contact:	**FIU Technical Contact:**
Vwani Roychowdhury, Ph.D. Engineering Building IV 420 Westwood Plaza, Box 951594 Los Angeles, CA 90095 vwani@ee.ucla.edu 310-206-4975	Mark A. Finlayson, Ph.D. ECS Building, Room 362 11200 S.W. 8th Street Miami, FL 33199 markaf@fiu.edu 305-348-7988

Administrative/business contacts:

UCLA Administrative Contact:	**FIU Administrative Contact:**
Julia Zhu Office of Contract and Grant Administration 11000 Kinross Avenue, Suite 211 Los Angeles, CA 90095 ocga5@research.ucla.edu 310-794-0155 (office) 310-943-1658 (fax)	Robert Gutierrez Assistant VP for Research 11200 S.W. 8th Street MARC Building, Room 430 Miami, FL 33199-0001 gutierrr@fiu.edu 305-348-2494 (office) 305-348-4117 (fax)

Period of Performance:

July 1, 2017 through June 30, 2018

Relation and Overlap with Currently Proposed Effort:

This project focuses on disinformation narratives in the health space, focusing on automatically detecting and understanding anti-vaccination conspiracy theories. The main output of this project has been new algorithms for detecting narratives, on which we will be building in this proposed effort as described above in the technical approach.

4.b. Pending Support: (b)(4) [redacted]

Title of Proposal and Summary: (b)(4) [redacted]

Source and amount of funding: (b)(4) [redacted]

Percentage effort devoted to each project: 67% summer effort for PI Mark Finlayson (2 months) in the first year; 50% summer effort (1.5 months) for each of the remaining four years.

Identity of Prime Offeror and complete list of subwards: FIU is the prime offeror, and there are no subcontracts

Technical and Administrative contacts:

Technical Contact:	**Administrative Contact:**
Mark A. Finlayson, Ph.D. ECS Building, Room 362 11200 S.W. 8th Street Miami, FL 33199 markaf@fiu.edu 305-348-7988	Robert Gutierrez Assistant VP for Research 11200 S.W. 8th Street MARC Building, Room 430 Miami, FL 33199-0001 gutierrr@fiu.edu 305-348-2494 (office) 305-348-4117 (fax)

Proposed Period of Performance:

January 1, 2018 through June 30, 2022

Relation and Overlap with Currently Proposed Effort:

(b)(4) [redacted].

4.c. Pending Support: This Proposal

Title of Proposal and Summary: Detecting and Aligning Narrative Variants for Countering Disinformation

Source and amount of funding: ONR

	Base (1 year)	**Option (1 year)**	**Total (2 years)**
Direct	$149,262	$142,901	$292,163
Total	$200,000	$200,000	$400,000

Percentage effort devoted to each project: 33% summer effort for PI Mark Finlayson in the first year (1 month), 50% summer effort (1.5 months) in the second year.

Identity of Prime Offeror and complete list of subwards: FIU is the prime offeror, and there are no subcontracts

Technical and Administrative contacts:

Technical Contact:	**Administrative Contact:**
Mark A. Finlayson, Ph.D. ECS Building, Room 362 11200 S.W. 8th Street Miami, FL 33199 markaf@fiu.edu 305-348-7988	Robert Gutierrez Assistant VP for Research 11200 S.W. 8th Street MARC Building, Room 430 Miami, FL 33199-0001 gutierrr@fiu.edu 305-348-2494 (office) 305-348-4117 (fax)

Period of Performance:

September 1, 2017 to August 31, 2019

5. Qualifications

5.a. Ability to Perform the Research

Dr. Mark A. Finlayson is the single named key personnel on the project, and will be responsible for supervising all aspects of the research project. Dr. Finlayson is Assistant Professor of Computer Science in the School of Computing and Information Sciences (SCIS) at Florida International University (FIU). He received his Ph.D. from MIT in 2012, and from 2012-2014 was a Research Scientist at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). He is an authority on the computational analysis of narrative. His research spans several areas, including computational linguistics and natural language processing, artificial intelligence, and cognitive science, and he has published in venues across those fields. He is co-founder and general chair of the International Workshop on Computational Models of Narrative (CMN), which is the premier venue for scientific, computational work addressing narrative. The workshop held its 7th meeting in Krakow, Poland in 2016 (CMN'16). He was guest Editor-in-Chief of 2014 special issue of Literary and Linguistic Computer (LLC, now the Journal of Digital Scholarship in the Humanities, or DSH), the journal of record for digital humanities, on the topic of computational models of narrative.

5.b. Ability to Comply with Grant Conditions

Dr. Finlayson will devote 1 summer months in the first year of the project (33% of the summer effort, ~16% of his total time), and an additional 1.5 summer months in the option year. This is compatible with the effort proposed in his pending proposals. In the event that Dr. Finlayson becomes over-committed (e.g., because of subsequent funding awards not in preparation at this time) he will reduce his time on the other proposed projects accordingly.

5.c. Performance History

During his time at MIT Dr. Finlayson participated in obtaining and successfully performing on over $2.1M in research grants from NSF, DARPA, and the NIH, and his doctoral work was additionally funded by ONR, AFOSR, and OSD. Dr. Finlayson has proposed or been senior personnel on three military funded contracts:

(1) *Defining and Demonstrating Capabilities for Experience-Based Narrative Memory* (2009), DARPA FA8750-10-1-0076 (Seedling). This grant funded work for identifying and developing new technologies for managing battlefield information overload, and exploring the use of narrative as a way of measuring and addressing cultural misunderstandings. Under this contract, Dr. Finlayson directed the construction of a deeply annotated corpus of Russian cultural tales [6], and developed the Analogical Story Merging algorithm for plot extraction [8], [9].

(2) *Jumpstarting and Focusing the N2 Program via a Deeply-Annotated Corpus of Narratives* (2011), DARPA D12AF00018 (Seedling). The N2 program sought to reveal the neurobiological bases of persuasive narrative, and in this context Dr. Finlayson was engaged with this Phase 0 contract to develop the Islamic Extremist dataset for use by performer teams [7].

(3) *Modeling and Shaping Narrative Influence* (2012), DARPA D12AP00210. In this main phase contract, Dr. Finlayson and Dr. Patrick Winston (also at MIT), developed new techniques for modeling and understanding persuasive stories. They developed new techniques for plot analysis. This contract also produced the pilot data work demonstrate that character extraction is feasible.

5.d. Record of Integrity and Business Ethics

Neither FIU nor Dr. Finlayson has any history of unethical business or research behavior.

5.e. Eligibility to Receive Award

FIU is fully eligible to receive and administer federal research grants and contracts designated for institutions of higher education.

5.f. Organizational Experience

Florida International University (FIU) is a Carnegie-designated R1 public research university located in Miami, FL. One of South Florida's anchor institutions, FIU is also one of the 25 largest universities in the nation, and among the 10 largest public universities, enrolling over 54,000 students. FIU is a Hispanic-Serving and Minority Institution (HSI & MI) with over 60% Hispanic and 20% African American enrollment. The university had $171M in research expenditures in FY 15-16, and performs research across the full spectrum of academic areas, including medicine, the arts, sciences, and engineering. The university has a business school, law school, and medical school.

The School of Computing and Information Sciences (SCIS) has over 35 research-focused faculty across all areas of computer science. In 2015, SCIS ranked 39th in the nation among research universities for amount of federal funding for computer science. SCIS's faculty has among them two IEEE Fellows, five NSF CAREER award winners, five IBM Faculty award winners, one ACM fellow, two ACM Distinguished Scientists, and an ACM Distinguished Educator. SCIS is also home to the NSF Industry-University Cooperative Research Center for Advanced Knowledge Enablement (IUCRC-CAKE), as well as a participant in the NSF CREST Center for Aquatic Chemistry and the Environment (CAChE).

The University and School have performed on numerous federal contracts, and have all necessary administrative and financial instructions in place, including property control systems, human resources management, conflict of interest disclosure, and safety programs applicable to the work.

6. Citations

[1] Cox, M. R., *Cinderella: Three Hundred and Forty-Five Variants*. London: Folklore Society, 1893.

[2] Eisenberg, J. D. and Finlayson, M. A., "Automatic Identification of Narrative Diegesis and Point of View," in *Proceedings of the 2nd Workshop on Computing News Storylines (CNewsStory 2016)*, 2016.

[3] Eisenberg, J. D., Yarlott, W. V. H., and Finlayson, M. A., "Comparing Extant Story Classifiers: Results & New Directions," in *Proceedings of the 7th International Workshop on Computational Models of Narrative (CMN'16)*, 2016.

[4] Fay, M. P., "Story Comparison via Simultaneous Matching and Alignment," in *Third Workshop on Computational Models of Narrative (CMN)*, 2012, pp. 100–104.

[5] Fillmore, C. J., Baker, C. F., and Sato, H., "The FrameNet Database and Software Tools," in *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, 2002, pp. 1157–1160.

[6] Finlayson, M. A., "ProppLearner: Deeply Annotating a Corpus of Russian Folktales to Enable the Machine Learning of a Russian Formalist Theory," *Digital Scholarship in the Humanities*, 2015.

[7] Finlayson, M. A., Halverson, J. R., and Corman, S. R., "The N2 Corpus: A Semantically Annotated Collection of Islamist Extremist Stories," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 2014, pp. 896–902.

[8] Finlayson, M. A., "Inferring Propp's Functions from Semantically-Annotated Text," *Journal of American Folklore, Special Issue on Computational Folkloristics*, **129**(511), pp. 53–75, 2016.

[9] Finlayson, M. A., "Learning Narrative Structure from Annotated Folktales," Doctoral Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2012.

[10] Hussain, M. N., Obadimu, A., Bandeli, K. K., Nooman, M., Al-khateeb, S., and Agarwal, N., "A Framework for Blog Data Collection: Challenges and Opportunities," in *The IARIA International Symposium on Designing, Validating, and Using Datasets (DATASETS 2017)*.

[11] Ji, H., Grishman, R., Chen, Z., and Gupta, P., "Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges," *Proc. Recent Advances in Natural Language Processing 2009*, pp. 166–172, 2009.

[12] Nimmo, B., "Anatomy of an info-war: How Russia's propaganda machine works, and how to counter it," 2015.

[13] Schneider, D., Matuszek, C., Shah, P., Kahlert, R., Baxter, D., Cabral, J., Witbrock, M., and Lenat, D., "Gathering and managing facts for intelligence analysis," in *Proceedings of the International Conference on Intelligence Analysis*, 2005.

[14] Uther, H.-J., *The Types of International Folktales: A Classification and Bibliography Based on the System of Antti Aarne and Stith Thompson*. Helsinki: Academia Scientarium Fennica, 2004.

[15] Wikipedia, "2014 Ukrainian Revolution," *Wikipedia, The Free Encyclopedia*, 2017. [Online]. Available:

https://en.wikipedia.org/w/index.php?title=2014_Ukrainian_revolution&oldid=781084354. [Accessed: 04-Jun-2017].

[16] Wikipedia, "Annexation of Crimea by the Russian Federation," *Wikipedia, The Free Encyclopedia*, 2017. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Annexation_of_Crimea_by_the_Russian_Federation&oldid=783420825. [Accessed: 04-Jun-2017].

[17] Zipes, J., *The Trials and Tribulations of Little Red Riding Hood*, 2nd ed. London: Routledge, 1993.